

THE TRANSMISSION OF INFORMATION—II

R. M. FANO

TECHNICAL REPORT NO. 149

FEBRUARY 6, 1950

RESEARCH LABORATORY OF ELECTRONICS
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Research Laboratory of Electronics

Technical Report No. 149

February 6, 1950

THE TRANSMISSION OF INFORMATION - II

R. M. Fano

Abstract

This report which is a continuation of Technical Report No. 65 deals with the transmission of information through a discrete channel disturbed by noise. It considers, in particular, the problem of reducing the equivocation by coding. The behavior of the per-unit equivocation as a function of the channel equivocation is computed in an approximate manner for a number of coding schemes of increasing complexity. The results indicate that the limiting condition of vanishing per-unit equivocation is approached rather slowly with increasing code complexity.

The basic theory underlying the process of transmitting information by means of discrete signals has been presented in Technical Report No. 65. The transmitter was considered as a device able to indicate successively, by means of a conventional signal, one choice out of a finite group of possible choices. It was assumed that the received signal was uniquely related to the transmitted signal; in other words, that the receiver could detect correctly at all times the choices indicated by the transmitter. It is hardly necessary to say that this ideal situation does not belong to the physical world. In practice the choices indicated to the receiver by the signal are not necessarily the same as those which the transmitter intended to indicate, and are not uniquely related to them. This situation arises because of the presence of some random process which modifies the signal in an unpredictable manner. We shall refer to any such random process as "noise" regardless of its exact physical nature. This report is devoted to the study of the process of communication by means of discrete signals in the presence of noise.

The first three sections of the report are almost identical with the second chapter of the notes written in the Spring of 1949 to accompany a series of lectures given by the author for the staff of the Raytheon Manufacturing Company. The material covered in these sections does not extend beyond the theory developed by N. Wiener (1) and C. Shannon (2). The last section presents some recent work done by the author on the noise-reduction properties of coding.

1. Equivocation and net information transmitted

Our first task will be to compute the average amount of information received in the presence of noise when the statistical characteristics of both signal and noise are known. We shall limit our analysis at first to the case of independent selections which are also affected independently by noise. Let x be a discrete random variable representing the choice selected at the transmitter and y be another discrete random variable representing the corresponding choice indicated to the receiver by the signal. The same number N of choices are available at both the transmitter and the receiver. The statistical nature of the signal generated by the transmitter is given as usual in terms of a set of N probabilities $p(x)$. The average amount of information generated at the transmitter is then (3)

$$H_x = - \sum_x p(x) \log_2 p(x) \quad . \quad (1)$$

This summation is extended, of course, to the N values of x representing the N choices, although this is not indicated in detail in order to simplify the notations.

The effect of noise on the received signal can be represented conveniently by a set of N^2 conditional probabilities $p(x;y)$. More precisely $p(x;y)$ is the probability that choice y is received when choice x is transmitted. It is convenient to define three other sets of probabilities related to the two just defined. The joint probability that x is transmitted and y is received is given by

$$p(xy) = p(x) p(x;y) \quad . \quad (2)$$

The probability $p(y)$ that y is received is given by

$$p(y) = \sum_x p(xy) = \sum_x p(x) p(x;y) \quad . \quad (3)$$

Finally, the conditional probability that x was transmitted when y is received is given by

$$p(y;x) = \frac{p(xy)}{p(y)} = \frac{p(x) p(x;y)}{\sum_x p(x) p(x;y)} \quad . \quad (4)$$

Suppose that a choice y is indicated by the received signal. How much information is conveyed, on the average, by this indication? The difficulty in answering this question lies in the fact that the reception of y does not provide any definite knowledge about x ; that is, about the choice indicated by the transmitter. In fact, any one of the choices could be the one actually selected. On the other hand, the reception of y obviously provides some knowledge about x , at least in a probability sense.

Consider, for instance, a signal consisting of pulses and empty spaces, a priori equally likely to occur. Suppose that, because of noise, a pulse has a probability a of being received as an empty space, and an empty space has the same probability a of being received as a pulse. If a pulse is received we know that the probability that a pulse was actually transmitted is $1 - a$, while a is the probability that the pulse received was originally an empty space. It is clear that if $a < 0.5$, some knowledge about the transmitted signal has been provided by the reception of the pulse, even if we still do not know whether a pulse or an empty space was transmitted.

The situation can be stated in more general terms. The a priori probability that x will be selected is $p(x)$. After y has been received the a posteriori probability that x was selected is $p(y;x)$ as given by Eq. 4.

The problem is thus reduced to the evaluation of the information provided by such a change of probability. To solve this problem we observe first that the average amount of information that would be conveyed by the correct indication of x is equal to H_x , as given by Eq. 1, when the $p(x)$ alone are known. If, on the contrary, the $p(y;x)$ (corresponding to a particular choice y indicated by the signal to the receiver) were also known, the average amount of information that would be conveyed by the correct indication of x becomes

$$H_x(y) = - \sum_x p(y;x) \log_2 p(y;x) \quad . \quad (5)$$

It follows that the additional information corresponding to the knowledge of the N probabilities $p(y;x)$, as a result of the reception of y , must be, on the average,

$$H_R(y) = H_x - H_x(y) \quad . \quad (6)$$

To obtain the overall average amount of information per indication received, we must average $H_R(y)$ with respect to y . The final result is then

$$H_R = \sum_y p(y) H_R(y) = H_x - H_{y;x} \quad (7)$$

where

$$H_{y;x} = - \sum_y p(y) \sum_x p(y;x) \log_2 p(y;x) \quad . \quad (8)$$

The quantity $H_{y;x}$ represents the additional average amount of information required for the determination of x when y is known or, in other words, the average amount of information lost because of noise. The name "equivocation" has been given by Shannon to this quantity to stress the fact that it is a measure of the ambiguity of the received signal.

The idea involved in the above evaluation of H_R can be made clearer by means of the following physical illustration. Consider a system consisting of a transmitter and a receiver which can communicate through a noiseless channel as well as through a noisy channel. Suppose that a long message has just been transmitted through the noisy channel, and suppose that the knowledge of the message received has somehow been fed back to the transmitter. Then consider separately all the instances in which a particular choice y has been received together with the corresponding choices x transmitted.

If the noiseless channel is to be used next to provide the receiver with a definite knowledge of these x , how many binary selections are

required, on the average (per N order selection) for this purpose? Since both receiver and transmitter know the set of probabilities $p(y;x)$ specified by the y received, a new code can be used designed to be optimum for this set of probabilities. The number of binary selections that must be transmitted on the average for a given y is clearly just equal to $H_x(y)$ as given by Eq. 5. Therefore $H_x(y)$ is the additional information that must be transmitted to provide a definite knowledge of x . Of course, the code to be used depends on y and the specification of the code is just the information provided by the reception of y . Averaging the number of binary selections required for all possible codes yields the value $H_{y;x}$ given by Eq. 8. If the indication of x had been transmitted originally through the noiseless channel, and a code optimum for the set of probabilities $p(x)$ had been used, the average number of binary selections required would have been equal to H_x . In conclusion, the reception of y is equivalent, on the average, to a number of binary selections (units of information) equal to H_R as given by Eq. 7.

The soundness of the above analysis from a physical point of view is considerably impaired by the assumption that the transmitter has a knowledge of the message received. This assumption is at the basis of the whole physical procedure, since a more efficient code can be used only if transmitter and receiver are in a position to agree on it. It does not seem possible to eliminate this assumption while still maintaining arbitrariness of the message. In other words, whenever random errors occur in the transmission of a message, the number of binary selections required, on the average, to correct such errors cannot be made smaller than the amount of information conveyed by the message, if the knowledge of the errors occurred is not available to the transmitter. Similarly any experienced teacher knows that efficient teaching requires some sort of feedback, as provided, for instance, by questions from the students or by the correction of quizzes and homework. As a matter of fact, a good teacher, after finding out what is in the minds of his students, again covers the whole subject that has not been clearly understood, devoting to each part of its explanation a time roughly proportional to how far the students are from correct understanding. This process is very similar to making the number of binary selections in the second transmission proportional to the logarithm of the reciprocal of the probability $p(y;x)$, as judged by the receiver on the basis of the first transmission. It will be shown later how the need for feedback can be eliminated by performing the transmission in such a way as to make the probability of the occurrence of errors approach zero. It will be possible to achieve this result by transmitting only certain types of messages, rather than arbitrary messages. Roughly speaking, any uncertainty

on the part of the receiver will be avoided by transmitting only as much information as can be received correctly through the given noisy channel. Before proceeding to the discussion of this question we wish to interpret from other points of view the results already obtained, in order to throw additional light on the problem.

Suppose for a moment that the noise is considered as useful information; that it is actually another signal, and that we wish to receive both signals. The average amount of information per pair of selections x and y (by which the noise is completely determined) can be expressed in the three equivalent forms

$$H_{xy} = H_x + H_{x;y} = H_y + H_{y;x} \quad (9)$$

where H_x and $H_{y;x}$ are given by Eqs. 5 and 8, and

$$H_{xy} = - \sum_x \sum_y p(xy) \log_2 p(xy) \quad (10)$$

$$H_{x;y} = - \sum_x \sum_y p(x) p(x;y) \log_2 p(x;y) \quad (11)$$

$$H_y = - \sum_y p(y) \log_2 p(y) \quad (12)$$

The equivalence of the three expressions in Eq. 9 is evident from a physical point of view, and can be proved without difficulty by substituting Eqs. 5, 8, 10, 11, 12 in Eq. 9. Considering again the noise as an undesired disturbance, the information received through a noisy channel, given by Eq. 7, can now be expressed in two other equivalent forms by means of Eq. 9.

$$H_R = H_x - H_{y;x} = H_y - H_{x;y} = H_x + H_y - H_{xy} \quad (13)$$

We have already seen that the first form can be interpreted as the total information desired minus the additional information that must be transmitted after y has been received. The second form can be interpreted as the information corresponding to the reception of y if the channel were noiseless, minus that part of this information which is actually just noise. Finally, the third expression can be interpreted as the information corresponding to the reception of y if the channel were noiseless plus the information conveyed by x if y were not known, minus the information corresponding to the pair xy . In other words, H_R is the information common to x and y .

Further light can be thrown on these interpretations by considering

the limiting case of sequences of selections of length n approaching infinity. It is shown in Technical Report No. 65, Sect. 3 that, out of the ensemble of all possible sequences, we can form a group of sequences whose total probability approaches unity when n approaches infinity. The frequencies of occurrence of the different choices in the sequences of this group differ from the probabilities of the choices by amounts that also approach zero when n approaches infinity. The number of sequences in these groups is then equal to the reciprocal of the probability of each sequence, that is to

$$S_x(n) = \prod_{x=0}^{N-1} [p(x)]^{-np(x)} = 2^{nH_x} \quad (14)$$

The same is true, of course, if we write y instead of x . Similarly, the number of sequences of errors produced by noise belonging to the group whose total probability approaches unity when n approaches infinity is

$$S_{x;y}(n) = 2^{nH_{x;y}} \quad (15)$$

This is, of course, the number of different sequences of y -selections that have probability approaching unity of originating from a given sequence of x -selections. The corresponding number of sequences of x -selections that have a probability approaching unity of having caused a given sequence of y -selections is

$$S_{y;x}(n) = 2^{nH_{y;x}} \quad (16)$$

The significance of the groups of sequences defined above is indicated schematically in Fig. 1. It is clear that, when n is very large, we can neglect in our reasoning all sequences not belonging to the above groups. In fact, the number of binary selections required on the average to transmit or to correct the transmission of the sequences not belonging to these groups is negligible compared to the number required for the others, regardless of the code used. On this basis we can rewrite Eq. 7 in the form

$$H_R = \frac{1}{n} \log_2 \frac{S_x(n)}{S_{y;x}(n)} \quad (17)$$

and interpret it as $1/n$ of the amount of information provided by the specification of which group of $S_{y;x}(n)$ sequences of x -selections contains the transmitted sequence. The second expression for H_R in Eq. 13 can be interpreted as $1/n$ of the information provided by the reception of a y -sequence when it is known that $S_{x;y}(n)$ different y -sequences can be produced by the same x -sequence. The third expression for H_R is a measure of the dependence

of the y-sequences on the x-sequences. It is $1/n$ of the difference between the number of binary selections required to specify both an x-sequence and a y-sequence if they were independent, and the number required to specify one

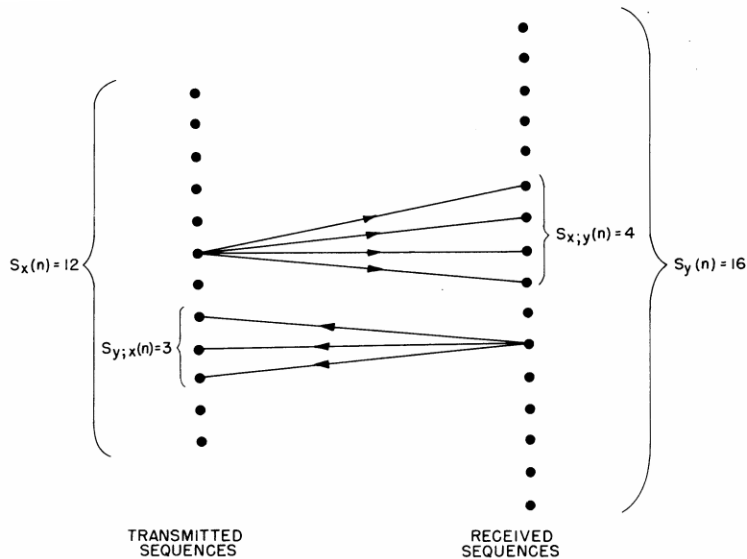


Fig. 1. Illustration of the limiting case of transmission through noise.

of the possible pairs of x and y-sequences. If there were no noise we would have

$$H_x = H_y = H_{xy} \quad (18)$$

because a y-sequence would be completely specified by the corresponding x-sequence.

The above analysis was limited to the case of sequences of independent selections. The corresponding results for the general case of non-independent selections would have more complex mathematical forms, but their significance would not change materially. Equations 9 and 13 are still valid in the general case, provided the proper expressions are used for H_x , H_y , $H_{y;x}$, $H_{x;y}$ and H_{xy} . The amounts of information represented by H_x , H_y and $H_{y;x}$ are never increased by a correlation between successive x-selections, as shown in Sect. 5 of Technical Report No. 65. It will be remembered that H_x represents the information conveyed by the knowledge of x, H_y represents the information (about y) conveyed by the knowledge of y, and $H_{y;x}$ represents the information conveyed by the knowledge of x when y is already known. $H_{x;y}$ represents the information (about y) conveyed by the knowledge of y when x is already known, and thus depends only on the noise and on the first probabilities $p(x)$. It follows that H_R , the information about x resulting from the knowledge of y, is never increased by any correlation

between successive x-selections, as indicated by Eq. 13. This general result is important in the following determination of the capacity of a given channel.

2. Channel capacity

The preceding section has been devoted to the evaluation of the information received in the presence of noise, when the statistical characters of the signal transmitted as well as of noise are specified. We now wish to determine the maximum rate at which information can be received through a channel able to handle a specified number N of choices and disturbed by a noise whose statistical character is also specified. In other words, we wish to determine the capacity of a channel having specified characteristics. Mathematically speaking, this problem amounts to the maximization of the quantity H_R with respect to the $p(x)$, subject to the constraint

$$\sum_x p(x) = 1 \quad . \quad (19)$$

The $p(x;y)$, which represent the characteristics of the noise, are held constant in this maximization process. The x-selections may be assumed to be independent, because, as shown above, H_R cannot be increased by any correlation between them. Using Lagrange's method, we equate to zero the partial derivatives of the function

$$H_R + \lambda \sum_x p(x) = H_y - H_{x;y} + \lambda \sum_x p(x) \quad . \quad (20)$$

The constant λ is the indeterminate multiplier. Taking the partial derivatives of Eq. 20 with respect to the $p(x)$ we obtain a set of equations, one per value of x , of the form

$$-\sum_y p(x;y) [1 + \ln p(y)] + \sum_y p(x;y) \ln p(x;y) + \lambda \ln 2 = 0 \quad . \quad (21)$$

These equations are more conveniently written in the form

$$-\sum_y p(x;y) \log_2 p(y) = -C + H_y(x) \quad (22)$$

where

$$C = -\lambda + \frac{1}{\ln 2} \sum_y p(x;y) = \frac{1}{\ln 2} - \lambda \quad (23)$$

and

$$H_y(x) = - \sum_y p(x;y) \log_2 p(x;y) \quad .$$

We observe first that C is the capacity of the channel, that is, the maximum value of H_R . By multiplying each term of Eq. 22 by $p(x)$ and adding for all values of x we obtain

$$- \sum_x \sum_y p(x) p(x;y) \log_2 p(y) - \sum_x p(x) H_y(x) = H_y - H_{x;y} = C \quad . \quad (24)$$

To determine the set of $p(x)$ that yields the maximum value of H_R , that is, C , it is convenient to write the set of Eqs. 22 in matrix form as follows

$$\left[P(x;y) \right] \cdot \log_2 p(y) = - (C + H_y(x)) \quad (25)$$

$\left[p(x;y) \right]$ is a square matrix representing the noise characteristics, in which the terms of any row correspond to the same x and the term of any column correspond to the same y . Solving this matrix equation for the column matrix $\log_2 p(y)$ yields

$$\log_2 p(y) = - \left[h(x;y) \right] \cdot (C + H_y(x)) \quad (26)$$

$\left[h(x;y) \right]$ is another square matrix in which the elements of each column are the cofactors of the elements of the corresponding row in $\left[p(x;y) \right]$ divided by the value of the determinant $\left| p(x;y) \right|$. In matrix notations

$$\left[h(x;y) \right] = \left[p(x;y) \right]^{-1} \quad . \quad (27)$$

For instance if both x and y have only two values, 0 and 1,

$$\begin{aligned} \left[p(x;y) \right] &= \begin{bmatrix} p(0;0) & p(0;1) \\ p(1;0) & p(1;1) \end{bmatrix} \\ \left[h(x;y) \right] &= \begin{bmatrix} h(0;0) & h(1;0) \\ h(0;1) & h(1;1) \end{bmatrix} \end{aligned} \quad (29)$$

where, for instance,

$$h(0,1) = \frac{-p(1;0)}{p(0;0)p(1;1) - p(0;1)p(1;0)} \quad . \quad (30)$$

We can now express $p(y)$ in the form

$$p(y) = 2^{- \sum_x h(x;y) [C + H_y(x)]} \quad . \quad (31)$$

In order to express $p(x)$ as a function of the $p(y)$ and the $p(x;y)$, we observe that

$$p(y) = \sum_x p(x) p(x;y) \quad (32)$$

or, in matrix notations,

$$p(y) = \left[p(x;y) \right]_t \cdot p(x) \quad (33)$$

where $\left[p(x;y) \right]_t$ is obtained from $\left[p(x;y) \right]$ by interchanging rows with columns. We have then

$$p(x) = \left[h(x;y) \right]_t \cdot p(y) \quad (34)$$

or

$$p(x) = \sum_y h(x;y) p(y) \quad (35)$$

Finally, the $p(x)$ corresponding to the maximum value of H_R , that is, to C , are given by

$$p(x) = \sum_y h(x,y) 2^{-\sum_x h(x;y) [C + H_y(x)]} \quad (36)$$

On the other hand, the value of the capacity C must satisfy Eq. 19 or the equivalent constraint

$$\sum_y p(y) = 1 \quad (37)$$

Substituting Eq. 32 for the $p(y)$ yields

$$\sum_y 2^{-\sum_x h(x;y) [C + H_y(x)]} = 1 \quad (38)$$

from which C can be determined, all other quantities in this equation being known.

Because of the fact that Eq. 38 is a transcendental equation in the unknown C , it is difficult to visualize how C depends on the noise. Fortunately, all difficulties disappear if $H_y(x)$ is independent of x , that is, roughly speaking, if the noise has the same disturbing effect on all x . This happens, for instance, in the case of binary pulse transmission if the probability of a pulse being received as an empty space is equal to the

probability of an empty space of being received as a pulse. In any such case

$$H_{x;y} = \sum_x p(x) H_y(x) \quad (39)$$

is independent of the $p(x)$; that is $H_{x;y}$ depends only on the noise and not at all on the signal. It follows then from

$$H_R = H_y - H_{x;y} \quad (40)$$

that the maximum value of H_R is obtained when H_y is a maximum, that is, when all choices y are a priori equally likely to be received. If the number of choices is N we have then for the channel capacity

$$C = \log_2 N - H_{x;y} \quad (41)$$

and for the values of the $p(x)$

$$p(x) = \frac{1}{N} \sum_y h(x;y) \quad (42)$$

Considering as an example the case of binary pulse transmission mentioned above, and letting a be the probability of an empty space being received as a pulse and vice versa, we obtain

$$C = 1 + \left[a \log_2 a + (1-a) \log_2 (1-a) \right] \quad (43)$$

and

$$\begin{bmatrix} p(x;y) \end{bmatrix} = \begin{bmatrix} 1-a & a \\ a & 1-a \end{bmatrix} \quad (44)$$

$$\begin{bmatrix} h(x;y) \end{bmatrix} = \frac{1}{1-2a} \begin{bmatrix} 1-a & -a \\ -a & 1-a \end{bmatrix} \quad (45)$$

The $p(x)$ are

$$p(0) = p(1) = \frac{1}{2} \quad (46)$$

as expected, because of the symmetry of the problem. In general, the $p(x)$ are different, even when the $H_y(x)$ are equal, unless the sets of $p(x;y)$ for any given x consist of the same set of numbers permuted with respect to y .

For a noise corresponding to a probability of error $a = 0.01$, the channel capacity becomes approximately 0.9, that is 90 percent of its value in the absence of noise. It is interesting to note, in this regard, that

the probability of error is far from being a direct measure of the loss of information. In the most general case of arbitrary $H_y(x)$ the order of magnitude of the channel capacity can be quickly evaluated by noting that

$$\log_2 N - [H_y(x)]_{av} < C < \log_2 N - [H_y(x)]_{min} \quad (47)$$

where $[H_y(x)]_{min}$ is the smallest $H_y(x)$ of the set, and

$$[H_y(x)]_{av} = \frac{1}{N} \sum_x H_y(x) \quad (48)$$

3. Ideal transmission and the reduction of the equivocation.

The discussion presented in the preceding sections has led to the determination of a theoretical limit for the amount of information per selection that can be received in the presence of a given noise. This limit on the rate of transmission of information, however, does not provide a complete solution to the problem of interest, even from a theoretical point of view. In fact our problem is not simply to receive as much information as possible, but also to receive as much of the information transmitted as possible. In other words, in addition to knowing the upper limit for the information received, H_R , we would also like to know the lower limit of the ratio $H_{y;x}/H_x$, which represents the per-unit equivocation or per-unit information lost.

In order to clarify this point, suppose we wished to transmit independent selections, as in the case of a telegraphic message consisting of numbers rather than words. As long as H_R is different from zero, we could theoretically transmit correctly all the information desired using a minimum number of binary selections. This could be done by repeating over and over the same message, using each time a more efficient code as indicated in Sect. 2. The main difficulty arising in such a procedure, as pointed out also in Sect. 1, is that after each transmission the message received would have to be retransmitted back so that a more efficient code could be used in the successive transmissions. This method of transmission is likely to be impractical. What we would like to do instead is to perform only one transmission, using as many binary selections as necessary to insure, for any practical purposes, the correct reception of the message. It is clear that this can be done when the noise is sufficiently low, as in ordinary telegraph transmission. We would like to know whether or not this can be done in general as long as H_R is different from zero. In other words, we would like to know whether information can be transmitted in such a way as to decrease the equivocation produced by noise and, if so, what is the lower

theoretical limit for the equivocation. It will be shown below that the answer to this question is yes and that the theoretical limit for the equivocation is zero. In other words, the probability of incorrect reception of a message can be made as small as desired; and, the number of binary selections used in the transmission does not need to be larger, on the average, than the information content of the message divided by H_R . The price for this reduction of the equivocation is paid in terms of a more involved coding of the information, and a correspondingly longer delay.

The manner in which such a result can be achieved, or at least approached, is suggested by what people often do instinctively when communicating by means of speech or visual indications. Suppose two persons must exchange some information while in a place so noisy that normal conversation is impossible, and that they cannot see each other so that any visual form of communication is precluded. What is often done in such a case is to decide ahead of time that only a certain very restricted group of words will be used, the words being selected in such a way that they have the least chance of being confused with one another. More specifically, suppose a man in a factory must indicate to another man whom he cannot see when to start or stop a machine. He certainly will not use such words as start and stop but rather go and stop or any two other conventional sounds that differ considerably from each other. The man who receives the information would not, in general, be able to recognize the words spoken if he had to differentiate between all possible words instead of the restricted number agreed upon. In other words, the possibility of his receiving correctly the information transmitted is based on his ability to decide that the sound heard is more similar to one word of the group agreed upon than to any other word of the same group. The larger is the noise, the more different the words of the group must be and, therefore, the smaller the number of words in the group.

Let us adapt the idea involved in this human communication process to the case of communication by means of sequences of selections. It is clear that we must select out of the ensemble of all possible sequences of a given length a group of sequences which, when transmitted, have the least probability of being confused with one another. More precisely, suppose we select M sequences out of the ensemble of $S = N^n$ possible sequences of n N -order selections and we consider these M sequences as forming a new ensemble of choices from which to select. Let us indicate with ξ the transmitted sequence of this ensemble and with η the received sequence. It must be noted that, while the transmitted sequence belongs to the group of M selected, the received sequence may be any one of the S possible sequences of the same length. Each sequence ξ has a probability $p(\xi)$ of being

transmitted and a conditional probability $p(\xi; \eta)$ of being transformed into a particular sequence η by the noise. From these two sets of probabilities we can obtain the conditional probability

$$p(\eta; \xi) = \frac{p(\xi) p(\xi; \eta)}{\sum_{\xi} p(\xi) p(\xi; \eta)} \quad (49)$$

that sequence ξ was transmitted when sequence η is received. In terms of these probabilities, the average amount of information transmitted per sequence is

$$H_{\xi} = - \sum_{\xi} p(\xi) \log_2 p(\xi) \quad . \quad (50)$$

The equivocation, that is, the amount of information lost, is

$$H_{\eta; \xi} = - \sum_{\xi} \sum_{\eta} p(\eta) p(\eta; \xi) \log_2 p(\eta; \xi) \quad (51)$$

where

$$p(\eta) = \sum_{\xi} p(\xi) p(\xi; \eta) \quad . \quad (52)$$

The improvement obtained by limiting the number of sequences that are transmitted is measured by the decrease of the per-unit equivocation $E = H_{\eta; \xi} / H_{\xi}$. Of course E becomes equal to $H_{y; x} / H_x$ when all possible sequences are transmitted. It will be shown below that, in all cases, the per-unit equivocation can be made as small as desired by making the number n of N -order selections in each sequence sufficiently large. At the same time the amount of information actually transmitted can be made to approach in the limit the value nC where C is the capacity of the channel, determined in Sect. 2. In other words, it is possible in theory to design a perfect communication system in which, while using the full capacity of a noisy channel, all the information transmitted is received correctly; that is, the probability of incorrect reception is zero. This fundamental theorem was first stated and proved by Shannon (2).

It was pointed out in Sect. 2 that the problem of transmission through noise becomes much easier to handle in the limiting case when the length n of the sequences approaches infinity. In this case, in fact, the ensemble of sequences that may be formed at the transmitter can be divided in two groups, whose total probabilities of occurrence are one and zero

respectively. The group with probability one contains (see Eq. 14)

$$S_x(n) = 2^{nH_x}$$

sequences in all of which the different choices appear with frequencies equal to their probabilities, and which, therefore are equally probable. The same is true for the received sequences, in which case the group with unity probability contains

$$S_y(n) = 2^{nH_y}$$

equally probable sequences. Similarly, each transmitted sequence (belonging to the probable group) has unity probability of being transformed by noise into any one of a group of

$$S_{x;y}(n) = 2^{nH_x}$$

sequences which, in turn, are equally probable (see Eq. 15). Conversely when a particular sequence (belonging to the probable group) is received, the corresponding sequence transmitted belongs, with unity probability, to a group of

$$S_{y;x} = 2^{nH_{y;x}}$$

sequences which are also equally probable (see Eq. 16). This situation is summarized schematically in Fig. 1 in which each point represents an equally probable sequence and each line an equally probable transformation caused by noise.

Figure 1 indicates clearly that if the transmitted sequence is allowed to be any one of the $S_x(n)$ sequences belonging to the probable group, all that the receiver will be able to determine is a group of $S_{y;x}(n)$ sequences to which the transmitted sequence must belong. Suppose, however, that only a certain restricted group of M sequences are allowed to be transmitted. It might then be possible to select the sequences of this group in such a way that of the group of $S_{y;x}(n)$ sequences that the receiver determines only one sequence belongs to the restricted group. In this case no uncertainty would exist as to the sequence transmitted. This is the same as requiring that no two sequences of the M selected can be transformed by noise in the same sequence. Since each transmitted sequence may be transformed into $S_{x;y}(n)$ sequences, and since there are $S_y(n)$ different sequences that may be received, it must be

$$M \leq S_y(n)/S_{x;y}(n) = S_x(n)/S_{y;x}(n) = 2^{nH_R} \quad (53)$$

It will be shown below that the ratio

$$\frac{M/2^{nH_R}}{M/2^{nH_R}}$$

can be made as close to unity as desired. Unfortunately no direct way has been found to select the M-sequences in the optimum manner. All we can do, therefore, is to show that such an optimum selection must exist. This proof will be carried out by averaging (over all possible groups of M-sequences that can be formed) the probability that when a particular sequence is received there is an uncertainty as to the sequence transmitted. If this average probability is equal to p there exists at least one group of M-sequences for which the probability is smaller than p. We shall find that p can be made arbitrarily small as long as M is smaller than

$$\frac{2^{nH_R}}{2}$$

We have seen that the reception of a particular sequence specifies a group of $S_{y;x}(n)$ sequences to which the transmitted sequence must belong. Let us consider then all the possible groups of M-sequences that could have been formed and compute the probability p that, if the group used was selected at random, it contains at least one sequence common to the group of $S_{y;x}(n)$ sequences in addition to the one actually transmitted. Since one sequence of the M is specified, the remaining M - 1 may be any of the other $S_x(n) - 1$ sequences available at the transmitter. Thus the probability that a particular sequence is one of these M - 1 is

$$p_1 = \frac{M - 1}{S_x(n) - 1} \quad . \quad (54)$$

This is the same as the probability of drawing a black ball out of an urn containing $S_x(n) - 1$ balls of which M - 1 are black. The probability for two given sequences to be either one or both present among the M - 1 is

$$p_2 = \frac{M - 1}{S_x(n) - 1} + \left(1 - \frac{M - 1}{S_x(n) - 1}\right) \left(\frac{M - 1}{S_x(n) - 2}\right) = 2p_1 - \frac{(M - 1)(M - 2)}{(S_x(n) - 1)(S_x(n) - 2)} \quad . \quad (55)$$

It follows that

$$p_2 < 2p_1 \quad . \quad (56)$$

The reason for this inequality is that $2p_1$ contains twice the probability that both sequences will be present simultaneously. For the same reason we obtain in general

$$p_m < m p_1 \quad (57)$$

where p_m is the probability that any one or all of m given sequences are present among the $M - 1$. Therefore, in our particular case, the probability p that one or more of the sequences belonging to the group of $(S_{y;x}(n) - 1)$ will be present among the $M - 1$ sequences must be

$$p < \left(S_{y;x}(n) - 1 \right) \frac{M - 1}{S_x(n) - 1} < M \frac{S_{y;x}(n)}{S_x(n)} \quad (58)$$

Finally, using Eqs. 14, 16 and 17 we obtain

$$p < M 2^{-n H_R} \quad (59)$$

It will be recalled that the M sequences forming the selected group are equally likely to be transmitted so that any one of them would convey an amount of information equal to $\log_2 M$. Since each sequence consists of n selections from the original M choices the amount of information per choice is

$$H_T = \frac{1}{n} \log_2 M$$

The inequality of Eq. 59 can then be written in the form

$$p < 2^{-n(H_R - H_T)} \quad (61)$$

This equation shows that p can be made as small as desired by increasing n (the length of the sequences) as long as H_T is smaller than H_R . On the other hand p is the average (over all possible groups of M -sequences) probability that when a particular sequence is received a doubt will exist as to the sequence transmitted. Therefore, there must be at least one group of M -sequences for which this probability can be made as small as desired, provided $H_T < H_R$. Using the symbolism introduced earlier in this section we have, when n approaches infinity and $H_T < H_R$

$$p(\xi) = 2^{-n H_T}; \quad H_\xi = n H_T \quad (62)$$

$$p(\eta;) = \begin{cases} 1 \\ 0 \end{cases} ; \quad H_{\eta;\xi} = 0 \quad (63)$$

Clearly, the per-unit equivocation vanishes when n approaches infinity provided $(H_R - H_T) > \epsilon$ where ϵ is an arbitrarily small positive constant. The maximum value that H_R can assume is the capacity C of the channel. Therefore, it is possible in theory to design a communication system which transmits information at a rate as close as desired to the capacity of the channel available, with a vanishingly small per-unit equivocation.

The situation can be summarized as follows. The statistical

characteristics of the channel noise specify a set of transmitter choice probabilities $p(x)$, for which H_R takes its maximum value C . If the number n of selections is sufficiently large, the problem reduces to that of $S_x(n)$ equally probable ξ -sequences, each of which may be transformed by noise into any one of a group of $S_{x;y}(n)$ η -sequences with equal probabilities. It is possible then to select approximately 2^{nC} ξ -sequences out of the $S_x(n)$ in such a way that the corresponding groups of $S_{x;y}(n)$ η -sequences do not have any sequences in common. If only these 2^{nC} sequences are allowed to be transmitted, no uncertainty will exist at the receiver as to the sequence transmitted each time. By coding information in these sequences in such a way as to make them equally likely to be transmitted, the amount of information transmitted and received per sequence can be made as close as desired to nC .

4. Coding for reduction of the per-unit equivocation.

It is hardly necessary to say that the ideal performance discussed above cannot be obtained in practice because it would require an extremely complex coding operation. The messages to be transmitted would have to be coded before transmission into very long (theoretically infinite) sequences of selections which in turn would have to be decoded after reception. The implications of this requirement go beyond the obvious practical considerations of equipment complexity. In fact no message or part of a message can be known precisely at the receiver until a whole sequence has been received and decoded. This does not mean that information is not received all the time, but rather that the information being received specifies more and more closely all the intelligence coded in the sequence. For instance, the noise disturbance may be such that a considerable amount of uncertainty concerning the first few selections of a sequence may exist until the last few selections are received. If we assume that the time required by the coding, transmission and decoding of a sequence is proportional to the length of the sequence itself, we arrive at the conclusion that any decrease of per-unit equivocation costs an additional time delay in the overall transmission process. In the limit, zero equivocation requires the use of infinitely long sequences which in turn implies an infinite time delay. It is natural to ask in such a situation what is the minimum per-unit equivocation that can be obtained with a given channel for specified length of sequences and specified rate of transmission of information. In symbols, what is the minimum value of E that can be obtained for given n , H_ξ/n and channel characteristics? Unfortunately this question has not yet been answered, although it is reasonable to expect that the larger the difference $C - H_\xi/n$, the smaller n needs to be for a given per-unit equivocation.

Another important consideration can be made regarding the performance of the communication system discussed above. Suppose that n has been made sufficiently large so that the per-unit equivocation is negligibly small for $(H_{\xi}/n - C_0) < \epsilon$, where ϵ is a very small number. Then consider the behavior of the per-unit equivocation when the capacity of the channel C is varied by changing the noise. The per-unit equivocation is negligible for any value of C larger than C_0 , but when C becomes smaller than C_0 , the per-unit equivocation must be

$$E \geq 1 - \frac{C}{C_0} \quad (64)$$

since C is by definition the maximum rate at which information can be received. In other words, the slope of the curve of the per-unit equivocation as a function of C/C_0 is discontinuous at the point $C/C_0 = 1$. This is essentially the threshold phenomenon common to all noise-reducing systems such as frequency modulation, pulse code modulation, etc. The behavior of the per-unit equivocation as a function of $H_{y;x}$ is illustrated in Fig. 2 for

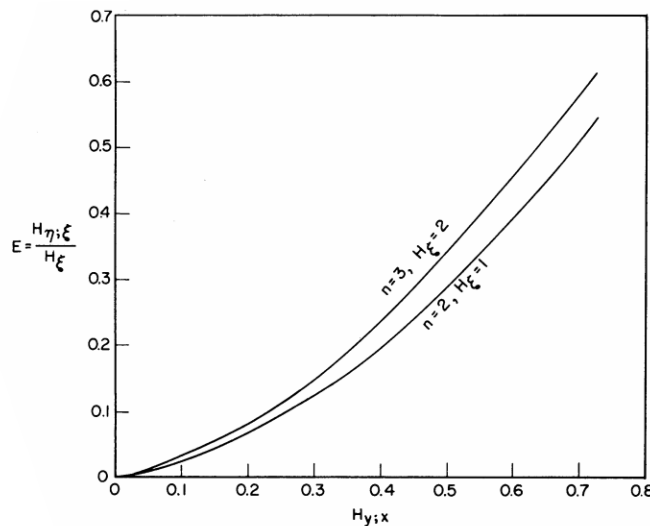


Fig. 2. The reduction of per-unit equivocation in two simple cases.

two typical coding schemes. The system considered employs binary equally likely selections and the noise is assumed to be such that either choice has a probability p of not being transformed into the other choice by noise and, of course, $q = 1 - p$ of being transformed. In the case of $n = 2$ the two sequences 00 and 11 constitute an optimum set of $M = 2$ equally probable ξ -sequences in which case $H_{\xi} = 1$. One obtains for H_{η} , $H_{\xi;\eta}$ and $H_{\eta;\xi}$

$$H_{\eta} = - \sum_{\eta} p(\eta) \log_2 p(\eta) = - \left[2pq \log_2 2pq + (1 - 2pq) \log_2 (1 - 2pq) - 1 \right] \quad (65)$$

$$H_{\xi;\eta} = 2H_{x;y} = -2(p \log_2 p + q \log_2 q) \quad (66)$$

$$H_{\eta;\xi} = 1 - (H_{\eta} - H_{\xi;\eta}) \quad (67)$$

In the case of $n = 3$ the sequences 000, 011, 110, 101 constitute an optimum set of M equally probable ξ -sequences, for which $H_{\xi} = 2$. We obtain for H_{η} , $H_{\xi;\eta}$ and $H_{\eta;\xi}$

$$H_{\eta} = - \left\{ (q^3 + 3p^2q) \log_2(p^3 + 3p^2q) + \left[1 - (q^3 + 3p^2q) \right] \log_2 \left[1 - (q^3 + 3p^2q) \right] - 2 \right\} \quad (68)$$

$$H_{\xi;\eta} = 3H_{x;y} = -3(p \log_2 p + q \log_2 q) \quad (69)$$

$$H_{\eta;\xi} = 2 - (H_{\eta} - H_{\xi;\eta}) \quad (70)$$

Figure 2 shows for the two cases considered above the behavior of E , the per-unit equivocation, as a function of $H_{y;x}$, the equivocation for the individual binary selections which is a measure of the noise present in the system. It will be observed that both curves show the beginning of a threshold phenomenon, in the sense that the slope increases with $H_{y;x}$. The reduction of per-unit equivocation is given by the difference between the ordinates of the curves.

The labor involved in considering in the same manner more complex coding schemes becomes increasingly prohibitive as the value of n increases. On the other hand it would seem worth while to determine, in an approximate fashion at least, how fast the per-unit equivocation decreases with increasing n for a constant rate of transmission of information. Similarly it would be very instructive to see how the threshold phenomenon becomes more pronounced with increasing n . Unfortunately, the behavior of the per-unit equivocation does not seem to be representable by means of a reasonably simple mathematical expression, even in an approximate manner. But it has been possible to compute a few representative curves for a typical case of possible practical importance.

Figure 3 illustrates the behavior of the per-unit equivocation after coding ($E = H_{\eta;\xi}/H_{\xi}$) as a function of the channel equivocation ($H_{y;x}$) for binary equally likely selections. Either binary choice has the same probability q of being received incorrectly. $H_{y;x} = 1 - C$ is a measure of the noise present in the channel. The figures on the curves, $n = 5, 10, 15, 20, 67$, indicate the length (number of binary selections) of the sequences used in the coding scheme. For all curves the rate of transmission H_{ξ}/n is kept constant and equal to $1/5$; that is, the transmission takes place at a

rate equal to one-fifth of the maximum rate that would be possible in the absence of noise. The straight line marked $n = \infty$ represents the theoretical boundary of the region over which the rate at which information is received

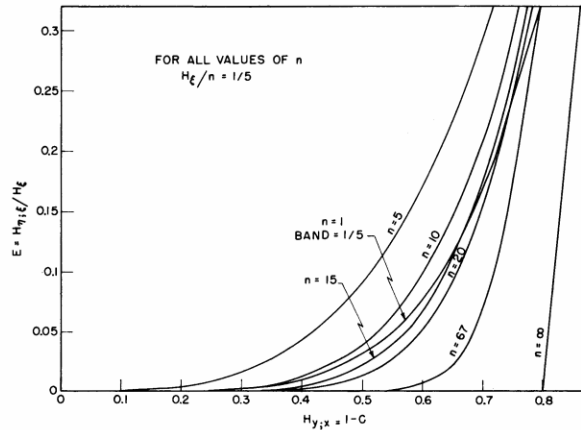


Fig. 3. Per-unit equivocation as a function of channel equivocation.

is smaller than the capacity of the channel, $1/5 (1 - E) \leq C$. The manner in which the curves have been computed is discussed below.

Let us consider sequences of n binary selections and think of them, for convenience, as numbers written in the binary system. Let q be the probability that any one digit be transmitted incorrectly; $p = 1 - q$ is then the probability that any one digit be transmitted correctly. The probability that a number ξ with n digits be received as a number η is given by

$$p^m q^{n-m}$$

where m is the number of digits that are the same in the two numbers. The number of η -numbers which have m digits in common with a particular number ξ is

$$\frac{n!}{(n-m)! m!} \quad .$$

It follows that when ξ is transmitted, the probability that an η is received which has in common with ξ less than k digits, is

$$Q_k = \sum_{m=0}^{m=k-1} \frac{n!}{(n-m)! m!} p^m q^{n-m} = q^n + nq^{n-1} p + \frac{n(n-1)}{2!} q^{n-2} p^2 + \dots + \frac{n(n-1)\dots(n-k)}{(k-1)!} q^{n-k+1} p^{k-1} \quad . \quad (71)$$

This expression is easily recognized as a partial sum of the terms of the

binomial $(q + p)^n = 1$. This partial sum can be expressed in integral form as

$$Q_k = I_q(n - k + 1, k) = \frac{\int_0^q x^{n-k} (1-x)^{k-1} dx}{\int_0^1 x^{n-k} (1-x)^{k-1} dx} \quad (72)$$

The definite integral at the denominator is known as the beta-function and $I_q(r, k)$ is known as the normalized incomplete beta-function. In our case $r = n - k + 1$ and $r + k = n + 1$. The identity of $I_q(r, k)$ with the partial sum of Eq. 71 can be easily checked by integrating the numerator of Eq. 72 by parts. An extensive tabulation of $I_q(r, k)$ is available (4) for integral values of r and k up to 50, and for values of q in steps of 0.01. Tables of q for given values of $I_q(r, k)$ and series expressions for computing $I_q(r, k)$ are also available in the literature (5,6).

Returning to our problem, we certainly can select M_ξ -numbers such that no η -number will have in common k or more digits with any two of the selected numbers. Then, if only these M_ξ -numbers are used by the transmitter, the probability of incorrect reception will be smaller than Q_k . It remains to be determined which is the largest value of M for which this condition is satisfied. The number of η -numbers which have in common k or more digits with a given ξ -number is given by

$$2^n I_{\frac{1}{2}}(k, r) = \sum_{m=0}^{n-k} \frac{n!}{(n-m)! m!} = n + \frac{n(n-1)}{2!} + \dots + \frac{n(n-1)\dots(k+1)}{(n-k)!} \quad (73)$$

This number represents a fraction $I_{1/2}(k, r)$ of the total number (2^n) of η -numbers. It follows that the largest value of M for which the probability of error is smaller than Q_k cannot be larger than $(I_{1/2}(k, r))^{-1}$

$$M \leq \left[I_{\frac{1}{2}}(k, r) \right]^{-1} \quad (74)$$

It does not follow necessarily that Eq. 74 can be satisfied with the equality sign, since we are not sure that we can find $[I_{1/2}(k, r)]^{-1}$ ξ -numbers which satisfy our requirement. On the other hand since it seems reasonable to expect that M can be made very close to $[I_{1/2}(k, r)]^{-1}$, at least for large values of n , we shall assume for our work that Eq. 74 holds with the equality sign. Thus, the amount of information that may be transmitted by means of an n -digit binary number with an overall probability of error equal to Q_k is

$$H_{\xi} = \log_2 M = - \log_2 \frac{I_1}{2}(k, r) \quad (75)$$

It has been pointed out above that the probability of error Q_k cannot be used directly as a measure of the loss of information resulting from the presence of noise. The quantity to be computed for this purpose is the per-unit equivocation, or in symbols

$$E = \frac{H_{\eta;\xi}}{H_{\xi}} = \frac{- \sum_{\eta} \sum_{\xi} p(\eta) p(\eta;\xi) \log_2 p(\eta;\xi)}{\log_2 M} \quad (76)$$

where the summation over ξ is limited to the M_{ξ} -numbers selected. The computation of the numerator of Eq. 76 presents a serious problem. The author could not find a way of performing such a computation which would avoid the laborious determination of the optimum set of M_{ξ} -numbers. A satisfactory approximate computation can be carried out, as follows.

Let us consider the η -numbers as divided into $M = 2^{H_{\xi}}$ groups, each one corresponding to the M_{ξ} -numbers that are transmitted with equal probabilities. When a number η is received, the receiver interprets it as the ξ corresponding to the group to which η belongs. The probability that an error results from this procedure is Q_k as given by Eq. 72. Let us assume, because of lack of more precise knowledge, that, when the ξ transmitted is not the one corresponding to the group to which η belongs, any other ξ is just as likely to have been the number transmitted; that is, let us assume that $p(\eta;\xi)$ is equal to $1 - Q_k$ for the ξ corresponding to the group to which η belongs, and is equal to $Q_k/(M - 1)$ for any one of the other $M - 1$ numbers ξ . We are now in a position to express E in terms of known quantities as

$$E = \frac{- \sum_{\eta} p(\eta) \left[(1 - Q_k) \log_2 (1 - Q_k) + (M - 1) \frac{Q_k}{M - 1} \log_2 \frac{Q_k}{M - 1} \right]}{\log_2 M}$$

$$= - \frac{(1 - Q_k) \log_2 (1 - Q_k) + Q_k \log_2 \frac{Q_k}{M - 1}}{\log_2 M} \quad (77)$$

For values of Q_k smaller than about 0.1, that is, for values of Q_k of practical interest, we have, to a good approximation,

$$E \approx Q_k \frac{1 - Q_k + \ln \frac{1}{Q_k} + \ln (M - 1)}{\ln M} \quad (78)$$

where $\ln$ indicates the natural logarithm.

The physical significance of the approximations involved in the above computation of E can be interpreted as follows. Suppose that, after receiving a long message, the receiver had sent back to the transmitter the numbers received for checking purposes. What fraction of the whole information content of the message would have to be retransmitted to correct the errors? In the first place, the transmitter would have to indicate where, in the message, the errors have occurred. Since the probability of error is Q_k , the average number of binary digits (per number in the message) required to indicate whether each number is correct or incorrect is

$$- [Q_k \log_2 Q_k + (1 - Q_k) \log_2 (1 - Q_k)] \quad . \quad (79)$$

After the location of each error has been indicated, it is still necessary to retransmit the correct number; the numbers to be retransmitted constitute a fraction Q_k of the whole message. Each number may have $M - 1$ equally likely values, since the number received previously is known to be incorrect. Thus, this retransmission corresponds to an amount of information (per number in the original message) equal to

$$Q_k \log_2 (M - 1) \quad . \quad (80)$$

Dividing the sum of 79 and 80 by the information corresponding to each number of the original message ($\log_2 M$) yields the value of E given in Eq. 77. Thus, since an additional amount of information equal to $E \log_2 M$ (as given in Eq. 77) would be sufficient to specify completely the message, this same amount must be equal or larger than the information lost because of noise. In other words, the actual value of E must be equal to or smaller than the value given by Eq. 77. The difference between the two values of E results from the information conveyed by the reception of any number η , in addition to the knowledge that the number belongs to a particular group of η -numbers. This additional information would result in a smaller value of expression 80, because the $M - 1$ numbers would no longer be equally likely to be transmitted. It is clear from Eq. 78 that this difference is significant only when $Q_k > 1/(M - 1)$. In addition, the value of E given by Eq. 77 is a better estimate of the amount of information required in practice to correct the errors. A very involved technique would be required to perform the correction by means of a number of binary digits corresponding to the actual value of E .

The main results of the computations carried out as indicated above are presented in Fig. 3. Figure 4 illustrates the behavior of the information transmitted per number, H_ζ , as a function of r , the maximum number of incorrect digits for which the number transmitted is still recognized correctly

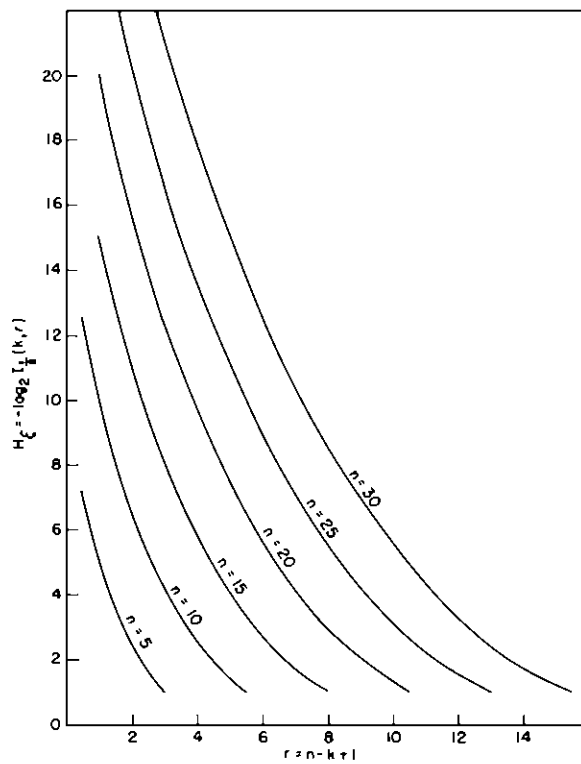


Fig. 4. Information transmitted as a function of maximum number of incorrect digits.

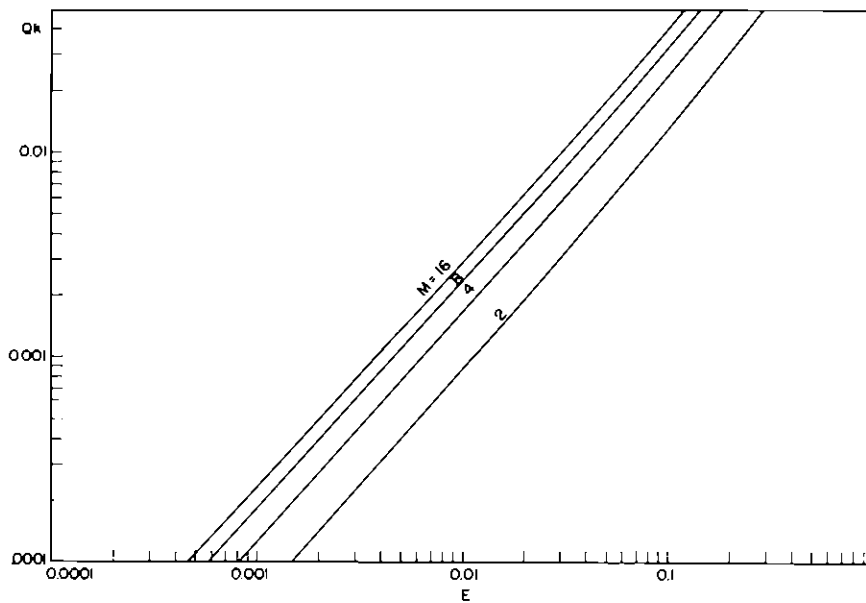
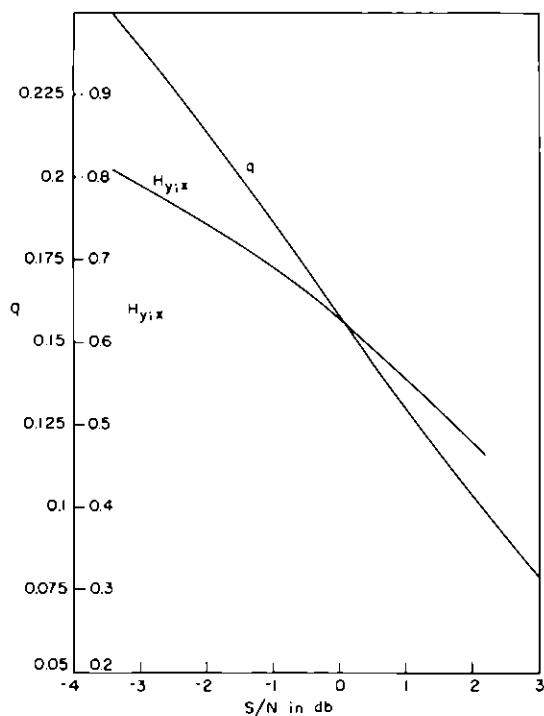
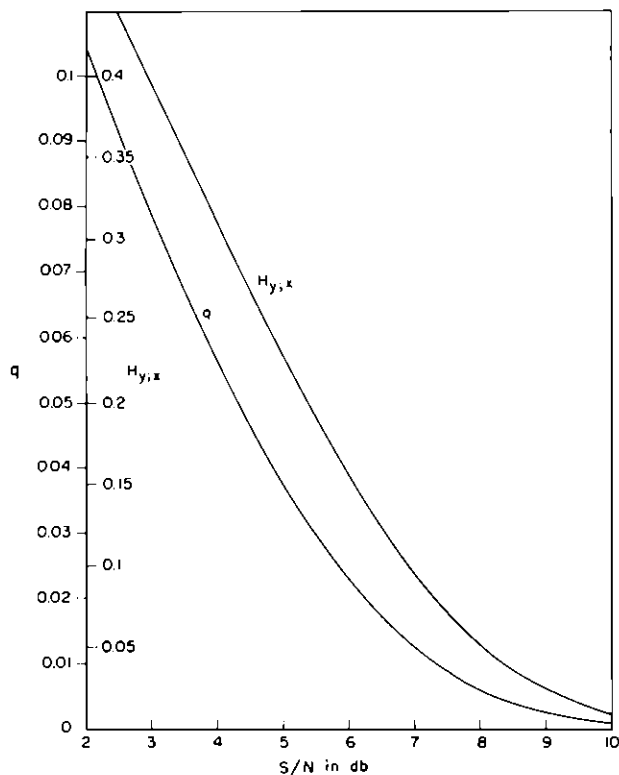


Fig. 5. Probability of error as a function of per-unit equivocation.



Figs. 6 and 7 Probability of error and channel equivocation as functions of signal-to-noise ratio.

by the receiver. It must be understood that these curves have been computed allowing fractional as well as integral values of r and k . In practice, of course, only integral values can be used, so that only discrete points of the curves correspond to physically realizable systems. Figure 5 presents a plot of the probability of error Q_k as a function of the per-unit equivocation E according to Eq. 78. Finally, Fig. 6 relates E and Q_k to the signal-to-noise ratio S/N in decibels for an unmodulated pulse system with two equally probable levels. The noise is assumed to be gaussian, that is, with an amplitude probability density

$$f(x) = \frac{1}{\sqrt{2\pi N}} e^{-\frac{x^2}{2N}} \quad (81)$$

where N is the noise power. The probability of error per binary pulse is given by

$$q = \frac{1}{\sqrt{2\pi N}} \int_{\sqrt{3}}^{\infty} e^{-\frac{x^2}{2N}} dx \quad (82)$$

where S is the average signal power, and the two pulse levels are equal to $\pm \sqrt{S}$. The same result is obtained, of course, for pulse levels equal to zero and $2\sqrt{S}$, if the d-c power is not taken into account.

The curves of Fig. 3 are drawn for a constant rate of transmission, that is for a constant value of $H_{\xi}/n = 1/5$. The parameter n is the number of digits (pulses) per ξ -number. The $H_{y;x}$ axis up to 0.8 together with the straight line $n = \infty$ represent the limiting case of the set of curves when n approaches infinity. The point $H_{y;x} = 0.8$ corresponds to the largest noise ($q = 0.245$, $S/N = -3.25$ db) for which 0.2 unit of information per pulse may be received; in other words, it corresponds to a channel capacity $C = 0.2$.

It is clear that this limiting behavior is approached rather slowly when n is increased. For $n = 67$ (the largest values for which tables of $I_q(r,k)$ are available) the curve has a marked threshold behavior but still it is not close to the theoretical limit.

It did not appear worth while to draw similar curves for larger values of H_{ξ}/n because they would approach the theoretical limit at an even slower rate. In fact the reduction of E is obtained by making use of the fact that it is unlikely that more than a certain fraction of digits are received incorrectly. The threshold phenomenon depends on how suddenly Q_k increases after q reaches a certain value and may be measured roughly by the ratio of the mean to the standard deviation of q for the density function dQ_k/dq . This ratio is given (6) by

$$T = \bar{q} / \sqrt{q^2 - (\bar{q})^2} = \sqrt{\frac{r}{k}} (r + k + 1) \quad . \quad (83)$$

Thus, for a constant value of $k + r = n + 1$, T is proportional to $\sqrt{r/k}$, which, in turn, decreases when k increases. On the other hand H_{ξ}/n increases with $k = (n + 1 - r)$ for a given n as shown in Fig. 4. We may conclude, therefore, that the threshold phenomenon, characteristic of a reduction of E becomes less pronounced, for a given n , when the value of H_{ξ}/n increases. It is interesting to note, in addition, that if we take T as a measure of the sharpness of threshold, and we observe that the ratio r/k does not change much for a given H_{ξ}/n , the sharpness of threshold is proportional to $\sqrt{n + 2}$.

A last point remains to be discussed in connection with Fig. 3. The curve for $n = 5$ corresponds to the case in which the per-unit equivocation is reduced by simply repeating the same binary digit 5 times. From a pulse transmission point of view this repetition is equivalent to increasing the pulse length by a factor of 5. This, in turn, would permit a 5 to 1 reduction of the bandwidth required for transmission and a consequent reduction

of the noise power by the same factor. In spite of this apparent equivalence, the process of repeating 5 times the same pulse indication and the process of reducing the bandwidth by a factor of 5 do not lead to the same value of per-unit equivocation. The curve marked "1/5 band" in Fig. 3 which applies to the latter process, was computed by means of the curves of Fig. 6. This curve lies considerably below the curve for $n = 5$ and mostly between the curves for $n = 10$ and $n = 15$. We must keep in mind, in this regard, that the values of E computed by means of Eq. 77 are somewhat higher than the correct values. Still, we should expect repetition to be less effective than band reduction in lowering the per-unit equivocation, because filtering is an averaging process which takes into account the whole noise wave, while the repetition process in our scheme takes into account only how often the noise amplitude is larger than a certain value.*

In conclusion, it appears that slowing down the rate of transmission by reducing the bandwidth is the best way to lower the per-unit equivocation, E , in the case of a very noisy channel. Larger reductions of E are possible, in theory, but they require very complex coding and appreciable transmission delay.

* Two coding schemes which take into account the whole noise wave are analyzed in a recent paper (7) by S. O. Rice.

References

1. N. Wiener: *Cybernetics* (John Wiley, N.Y. 1948).
2. C. E. Shannon: *A Mathematical Theory of Communication*. B.S.T.J. 27, Nos. 3 and 4 (July and Oct. 1948).
3. R. M. Fano: *The Transmission of Information*. Technical Report No. 65, Research Laboratory of Electronics, M.I.T. (1949).
4. K. Pearson: *Tables of the Incomplete Beta-Function* (The University Press, Cambridge, England, 1934).
5. C. M. Thompson: *Tables of Percentage Points of the Incomplete Beta-Function*. *Biometrika* 32, 151 (Oct. 1941).
6. H. E. Soper: *The Numerical Evaluation of the Incomplete Beta-Function*. Tracts for Computers, No. VII (Cambridge University Press, London, 1921).
7. S. O. Rice: *Communication in the Presence of Noise-Probability of Error for Two Encoding Schemes*. B.S.T.J. 29, No. 1 (Jan. 1950).